J Antimicrob Chemother 2023; **78**: 2323–2334
https://doi.org/10.1093/jac/dkad240 Advance Access publication 7 August 2023

Journal of Antimicrobial Chemotherapy

Prevalence of HIV-1 drug resistance mutations in proviral DNA in the Swiss HIV Cohort Study, a retrospective study from 1995 to 2018

Bashkim Jaha[1]†, Corinne D. Schenkel[1]†, Lisa Jörimann[1,2], Michael Huber [2], Maryam Zaheri[2], Kathrin Neumann[1,2], Christine Leemann[1,2], Alexandra Calmy[3], Matthias Cavassini [4], Roger D. Kouyos [1,2], Huldrych F. Günthard [1,2]‡, and Karin J. Metzner [1,2]*‡; on behalf of the Swiss HIV Cohort Study§

[1]Department of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, 8091 Zurich, Switzerland; [2]Institute of Medical Virology, University of Zurich, 8057 Zurich, Switzerland; [3]Division of Infectious Diseases, University Hospital Geneva, University of Geneva, Geneva, Switzerland; [4]Division of Infectious Diseases, University Hospital Lausanne, University of Lausanne, Lausanne, Switzerland

*Corresponding author. E-mail: karin.metzner@usz.ch
†B.J. and C.D.S. contributed equally to this work.
‡H.F.G. and K.J.M. contributed equally to this work.
§Members are listed in the Acknowledgements.

Received 26 April 2023; accepted 19 July 2023

Background: Genotypic resistance testing (GRT) is routinely performed upon diagnosis of HIV-1 infection or during virological failure using plasma viral RNA. An alternative source for GRT could be cellular HIV-1 DNA.

Objectives: A substantial number of participants in the Swiss HIV Cohort Study (SHCS) never received GRT. We applied a method that enables access to the near full-length proviral HIV-1 genome without requiring detectable viraemia.

Methods: Nine hundred and sixty-two PBMC specimens were received. Our two-step nested PCR protocol was applied to generate two overlapping long-range amplicons of the HIV-1 genome, sequenced by next-generation sequencing (NGS) and analysed by MinVar, a pipeline to detect drug resistance mutations (DRMs).

Results: Six hundred and eighty-one (70.8%) of the samples were successfully amplified, sequenced and analysed by MinVar. Only partial information of the *pol* gene was contained in 82/681 (12%), probably due to naturally occurring deletions in the proviral sequence. All common HIV-1 subtypes were successfully sequenced. We detected at least one major DRM at high frequency (≥15%) in 331/599 (55.3%) individuals. Excluding APOBEC-signature (G-to-A mutation) DRMs, 145/599 (24.2%) individuals carried at least one major DRM. RT-inhibitor DRMs were most prevalent. The experienced time on ART was significantly longer in DRM carriers ($P=0.001$) independent of inclusion or exclusion of APOBEC-signature DRMs.

Conclusions: We successfully applied a reliable and efficient method to analyse near full-length HIV-1 proviral DNA and investigated DRMs in individuals with undetectable or low viraemia. Additionally, our data underscore the need for new computational tools to exclude APOBEC-related hypermutated NGS sequence reads for reporting DRMs.

Introduction

After introduction of PIs and NNRTIs more than two decades ago, the era of combination ART (cART), incorporating several antiretroviral agents, has become state-of-the art treatment for people living with HIV-1 (PLWH), diminishing mortality and morbidity, due to its highly efficient viral suppression in PLWH.[1,2] Following the therapeutic innovations accompanied by systematic genotypic resistance testing (GRT), acquired and transmitted drug resistance (ADR and TDR, respectively) has declined in the following two decades in many resource-rich countries.[3–5]

Albeit major progress in ART, switching ART regimen in PLWH is often necessary to improve tolerability and toxicity, adherence, drug interactions or virological suppression.[6] For deciding if a

patient's ART needs modification, and how ART is modified, strategic clinical considerations of medical history, history of previous virological failure and GRT results are crucial, in order to prevent viral replication and subsequently achieve undetectable plasma viral loads, since ongoing viral replication is likely to cause DRMs under the selective pressure of ART.[7] Because emerging resistant viral strains can lead to TDR, hence endangering global therapeutic success, its monitoring on a global level remains imperative. Thus, generated resistance data do not only guide clinicians in decision-making, furthermore, they allow epidemiological analysis to understand viral evolution and spread of resistance, to develop strategies, and intervene on a public health level by adjusting first-line and prophylactic treatment. [8]

GRT in PLWH is recommended and routinely performed upon diagnosis of HIV infection or during episodes of viraemia following the initiation of ART in most resource-rich countries.[7,9] Usually, RT–PCR methods that require viral RNA templates, i.e. detectable viral load, are utilized to obtain genotypic sequence information of the RT and protease region of the HIV-1 genome.[7,9–11] More recently, GRT in proviral DNA has been suggested as an alternative to standard GRT methods in the absence of viraemia and/or states of low viraemia (<200 copies/mL plasma).[12–17]

In this retrospective study, we developed a near full-length HIV-1 next-generation sequencing (NGS) protocol and applied it to PBMC samples from 962 participants of the Swiss HIV Cohort Study (SHCS), allowing simultaneous GRT of the clinically relevant protease, RT and integrase independent of ongoing viral replication.

Materials and methods

Study population and design

All participants were members of the nationwide SHCS, who never underwent (protocoled) genotyping, neither at baseline nor at any other point during their participation in the SHCS. The most likely reason for unavailable sequence or GRT information was because there was either no available plasma sample from before ART or participants enrolled after initiation of ART. The SHCS is a scientifically multifaceted and versatile multicentre study founded in 1988 early in the HIV epidemic.[18] All PBMC samples were derived from specimens collected and stored in three participating centres in Switzerland, namely, University Hospital Geneva, University Hospital Lausanne and Clinique de La Source in Lausanne.

The SHCS has been approved by the ethics committee of the partaking institutions: Comité départemental d'éthique des spécialités médicales et de médicine communautaire et de premier recours, Hôpitaux Cantonale de Genève (number of approval: 01-142); Commission cantonale d'éthique de la recherche sur l'être humain, Canton de Vaud, Lausanne (number of approval: 131/01291012). All participants gave written informed consent.

PBMC samples and cellular DNA isolation

For each participant, the earliest available blood sample was requested from the SHCS biobanks, preferentially before participants had started ART. Specimens were delivered as PBMCs, either diluted DMSO or as dry cell pellets. Cellular DNA was isolated using the DNeasy Blood & Tissue Kit (QIAGEN) either manually, following the manufacturer's instructions, or with the QIAcube Connect (QIAGEN). DNA concentration was measured using the NanoDrop 1000 spectrophotometer (Thermo Scientific).

HIV-1 DNA amplification and NGS

We developed the final protocol, which contains two steps of enzymatic amplification, by integrating elements of the established RNA targeting protocols by Gall *et al.*[19] and Banin *et al.*[20] To allow high throughput, we incorporated our own primer modifications and optimized PCR conditions, such as magnesium concentration and cycling temperatures.

In the first step, one single near full-length HIV-1 genome, i.e. an ~9 kb amplicon, was amplified. The PCR reaction contained the 2× PrimeSTAR Max high-fidelity DNA polymerase (Takara Bio Inc.) premix (including reaction buffer, 2 mM Mg^{2+}, each dNTP at 0.4 mM), additional 1 mM $MgCl_2$, each primer at 0.3 μM (Table S1, available as Supplementary data at *JAC* Online) and 6 μL of isolated cellular DNA (representing 10% of the isolated cells) in a volume of 25 μL. The PCR was performed as follows: 98°C for 10 s; 35 cycles of 10 s at 98°C, 53°C for 15 s and 72°C for 4.5 min. Next, using 2 μL each of the first PCR reaction, two separate nested PCRs were performed to produce two overlapping amplicons. The elongation time was reduced to 2.5 min.

The two amplicons per sample were purified through solid-phase reversible immobilization using paramagnetic beads (AMPure XP kit, Beckman Coulter) and pooled at a concentration of 0.2 ng/μL. Library preparation was carried out using the Nextera XT DNA Library Preparation kit and Nextera Index kit (Illumina). Near full-length HIV-1 genomes were sequenced (2×151 bp paired-end reads) on the Illumina MiSeq platform using the MiSeq Reagent Kit v2 (300 cycles).[21]

Mutation and drug resistance analysis of HIV-1 proviral DNA

For systematic detection of drug resistance mutations (DRMs) in the HIV-1 *pol* region, we used the MinVar software (https://github.com/medvir/MinVar).[22] Briefly, using Illumina sequencing reads, it determines the closest HIV-1 subtype reference sequence from a local database and modifies it to construct the consensus sequence. It identifies amino acid mutations with respect to HIV-1 consensus B sequence in the Los Alamos database.[23] Using the Stanford HIV Drug Resistance Database algorithm,[24] MinVar ultimately generates a report about the HIV-1 subtype, DRMs with according frequencies, and potentially resulting drug resistance. We disregarded all mutations with a frequency of <15%.

Data synthesis and modification, i.e. filtration of APOBEC-signature DRMs, as well as statistical analysis of the results were performed using RStudio (R version 4.0.3). The filtering was done through removing all the relevant APOBEC-related amino acid mutations from the report after generation of the consensus sequence. All statistical tests were two-sided. Besides the base R functions,[25] we used functions for testing and visualization of the data from the following R packages: tidyverse,[26] tableone,[27] ggplot2[28] and ggpubr.[29]

Results

Characteristics of study population

We identified 1068 participants from the SHCS centres in Geneva and Lausanne, all without a record of a GRT. We received PBMC samples from 962 individuals. The median age of the participants, at the time of sample collection, was 40 years (Table 1). Three hundred and eighteen (30.1%) of the participants were female. Main ethnicities were white (610; 63.4%) and black (238; 24.7%). Homosexual (364; 37.9%) and heterosexual contact (HET, 429; 44.6%) accounted for the most common routes of transmission, followed by injection drug use (IDU, 82; 8.5%). Blood samples were collected between 1995 and 2018, with a median year of 2007. Eighty-seven (9.4%) participants were

Table 1. Characteristics of participants at the time of sampling stratified by outcome of HIV-1 DNA NGS analysis

			Sequencing			
Characteristic		All	Successful	Failed	*P* value	Test[a]
N		962	681	281		
Born in year, median (IQR)		1965 (1959–1975)	1965 (1958–1974)	1966 (1960–1976)	0.126	mw
Age (years), median (IQR)		40 (33–47)	40 (34–48)	38 (32–47)	0.063	*t*
Sex, *n* (%)	Male	644 (66.9)	461 (67.7)	183 (65.1)	0.487	χ^2
	Female	318 (33.1)	220 (32.3)	98 (34.9)		
Ethnicity, *n* (%)	White	610 (63.4)	438 (64.3)	172 (61.2)	0.287	χ^2
	Black	238 (24.7)	161 (23.6)	77 (27.4)		
	Hispano-American	34 (3.5)	20 (2.9)	14 (5)		
	Asian	27 (2.8)	20 (2.9)	7 (2.5)		
	Unknown	51 (5.3)	40 (5.9)	11 (3.9)		
	Other	2 (0.2)	2 (0.3)	0 (0)		
HIV-1 transmission group, *n* (%)	MSM	364 (37.8)	271 (39.8)	93 (33.1)	0.099	χ^2
	HET	429 (44.6)	293 (43)	136 (48.4)		
	IDU	82 (8.5)	61 (9)	21 (7.5)		
	Unclear (MSM, HET, IDU)	33 (3.4)	24 (3.5)	9 (3.2)		
	Blood products	9 (0.9)	3 (0.4)	6 (2.1)		
	Perinatal	16 (1.7)	9 (1.3)	7 (2.5)		
	Unknown	20 (2.1)	14 (2.1)	6 (2.1)		
	Other	9 (0.9)	6 (0.9)	3 (1.1)		
Year of sample collection, median (IQR)		2007 (2000, 2013)	2007 (2000, 2013)	2008 (1999, 2014)	0.622	mw
ART status (%)	On-ART	842 (90.6)	625 (93.3)	217 (83.8)	<0.001	χ^2
	Pre-ART	87 (9.4)	45 (6.7)	42 (16.2)		
CD4+ T cell count/mm^3 blood, median (IQR)		495 (325–709)	483 (314–678)	535 (348–757)	0.007	mw
CD4+ T cells (as % of lymphocytes), mean (SD)		26.7 (11.1)	25.6 (10.8)	29.4 (11.4)	<0.001	*t*
Log_{10} of HIV-1 RNA copies/mL plasma, median (IQR)		0 [0–1.7]	0 [0, 1.7]	0 [0, 1.4]	0.209	mw
Time on ART (years), median (IQR)		2.5 (0.8–6.2)	2.5 (0.9–6.4)	2.3 (0.4–5.6)	0.035	mw
Year of ART initiation, median (min–max)		2002 (1988–2018)	2001 (1989–2018)	2000 (1988–2018)		
HIV-1 subtype (%) (determined only in successfully sequenced cell samples)	A1		59 (8.6)			
	A2		2 (0.3)			
	B		443 (65.1)			
	C		38 (5.6)			
	CRF_11_cpx		3 (0.4)			
	CRF_14_BG		1 (0.1)			
	CRF01_AE		1 (0.1)			
	CRF02_AG		59 (8.6)			
	CRF06_cpx		4 (0.6)			
	CRF12_BF		2 (0.3)			
	D		5 (0.7)			
	F1		10 (1.5)			
	F2		3 (0.4)			
	G		24 (3.5)			
	Undetermined		27 (4)			
PBMC samples: isolated DNA conc.(ng/μL), median (IQR)		76.2 (42.1–124.3)	83.7 (47.2–131.7)	59.9 (31.6–95.8)	<0.001	mw

MSM, men who have sex with men; HET, heterosexual; IDU, injection/intravenous drug use; SD, standard deviation; min, minimum; max, maximum; conc., concentration.

[a]Applied statistical test of independence between the stratified groups with *t* for two-sample *t*-test, mw for Mann–Whitney *U*/Wilcoxon rank-sum, and χ^2 for Pearson's chi-squared. Percentages refer to the respective category only and not to the total amount of samples.

ART naive, whereas 842 (90.6%) had already experienced ART and therapy status was unknown in 33 cases. ART in ART-experienced participants was initiated between 1988 and 2018 (median 2002). The duration of ART experience ranged from 0 days to 26.3 years (median: 2.5 years). Viral load was detected in 279 (29%) participants (0.7–7.7 $\log_{10}$ HIV-1 RNA copies/mL plasma; median: 2.6 $\log_{10}$ copies/mL). Lowest CD4+ T cell count was 0 cells/mm^3 blood, with the highest amounting to 1855 cells/mm^3 (mean: 529 CD4$^+$ T cells/mm^3) (Table 1).

Factors associated with successful amplification and NGS of HIV-1 pol DNA

Amplification and NGS of amplicon 1 including HIV-1 *pol* were successful in 681/962 (70.8%) cases. We used MinVar to determine an average (IQR) coverage of 2535 (1540–3402) per nucleotide per sample. Of note, this is a result of 2×10^5 randomly drawn high-quality Illumina MiSeq reads per sample corresponding to a coverage of the order of 10^4.[22] Statistical analysis of participants' characteristics did not show any significant association between failure and success of sequencing regarding age, sex, ethnicity and HIV-1 transmission group (Table 1). Also, the age of the samples, i.e. the duration of sample storage, had no impact on the success (Table 1, Figure 1a).

We observed significant differences between successfully and non-successfully sequenced samples in the CD4+ T cell count/mm^3 ($P=0.007$, Table 1) and whether the participants had received ART or not ($P<0.001$). Specimens with fewer CD4+ T cells had a higher chance of successful amplification (Table 1, Figure 1b). In addition, we observed a higher proportion of ART-naive participants (pre-ART) in the samples that failed compared with those that succeeded in sequencing (Table 1). Analysing only this pre-ART group, the absolute count (per mm^3 of blood) of the CD4$^+$ T cells, as well as the viral load were compared regarding the outcome of amplification and NGS. We could hereby observe tendencies of lower CD4+ T cell count [414 (308–568) versus 492 (334–628) cells/mm^3] and higher viral load [4.2 (2.4–5) versus 3 (1.4–4.3) $\log_{10}$ HIV-1 RNA copies/mL plasma) in successfully sequenced samples but failed to provide significance, most likely due to small sample size (Table S2). Furthermore, missing sequencing information was strongly linked to a lower concentration of the isolated DNA, and thus to lower PCR input ($P<0.001$, Table 1). The median DNA isolation product in the successful amplification group was higher and amounted to 83.7 ng/μL (IQR 47.2–131.7) and 59.9 ng/μL (IQR 31.6–95.8) in the non-successful group (Table 1, Figure 1c).

Among the analysed sequences, HIV-1 subtype B was represented with the highest prevalence (443/681; 65.1%), followed by HIV-1 subtypes A1 with 8.6% (59/681), CRF02_AG with 8.6% (59/681) and C with 5.6% (38/681). Bioinformatic analysis did not assign a matching HIV-1 subtype in 27 samples (Table 1). Prevalence of DRMs in proviral DNA MinVar reports were received from 599/681 (88%) samples including DRMs, polymorphisms and ART-unrelated mutations in the three *pol* genes as compared with the reference HIV-1 genome. The sequence read coverage of the remaining 82 samples revealed large deletions in *pol*, as commonly observed in HIV-1 proviral DNA (Figure S1).[30,31] We investigated the prevalence of major DRMs that are listed in the Stanford HIV-1 database and reported by the IAS–USA.[24,32] DRMs with frequencies ≥15% in *pol* affecting PIs, NRTIs, NNRTIs and integrase strand transfer inhibitors (InSTIs) were included. At least one DRM was detected in 331/599 (55.3%) participants (Table 2). There were no significant trends of harbouring any DRMs in terms of age, sex, transmission group, HIV-1 subtype, CD4+ T cell count or viral load (Table 3). Following the prevalence over the years, a peak in DRM prevalence in 1995 is seen, with a nadir in 2005. Despite variations between selected years, there was neither a significant absolute increase nor decrease by time in prevalence of any major DRM (Figure 2; $P=0.651$). The results revealed an increased proportion of people with black or unknown ethnicity in the DRM carrier group (32.7% versus 24.3%, $P=0.019$). Further, the presence of a DRM was significantly associated with a longer time on ART, with DRM carriers being on ART for a median of 2.9 years (IQR 1.1–6.8) and non-carriers for 1.9 years (IQR 0.6–4.9) (Table 3).

The most abundant detected DRMs were: M46I (55/681; 7.8%) and D30N (53/681; 8.1%) for the protease; M230I (210/681; 30.8%), M184I (181/681; 26.6%), E138K (85/681; 12.5%) and D67N (47/681; 6.9%) for the RT; and E138K (50/681; 7.4%) and G140S (37/681; 5.4%) for the integrase region (Table 3). Of note, all those DRMs result from G-to-A mutations in the DNA sequence.

We filtered out the G-to-A DRMs from the dataset, more precisely only those DRMs associated with apolipoprotein B mRNA-editing enzyme catalytic polypeptide 3 (APOBEC3G/F) and listed in the Stanford HIV-1 database (Table 2).[33–35] The results were stratified in the same manner as before and are listed in Table 3. Of those initially labelled as DRM-carrying samples, 145/331 (43.8%) remained labelled as mutated. Again, we did not see any statistically significant differences between the participants still harbouring DRMs in age, sex, transmission group and HIV-1 subtype (Table 3). The prevalence course followed a similar shape, though on an approximately half-fold of the non-filtered values (Figure 2). Similarly, we found differences in the ethnicities ($P<0.001$) and time on ART ($P=0.02$; Table 3). Additionally, there was a significant difference found regarding the CD4+ T cell count [median 422.5 (IQR 252.8–650) in the DRM carriers versus 495 (IQR 332–681.5), $P=0.032$], its percentage of total lymphocytes [mean 23.9 (SD 11.8) in the DRM carriers versus 26.2 (SD 10.5), $P=0.026$] and in viral load [median 0 (IQR 0–2.3) in the DRM carriers versus 0 (IQR 0–1.5), $P=0.02$] (Table 3).

Prevalence of DRMs differs between the drug classes

After evaluation of characteristics of participants potentially contributing to the presence of any DRM, we analysed each drug class independently. RTI DRMs with abundance ≥15% were detected in 321/599 (53.7%) participants in the non-filtered and in 142/599 (23.7%) in the APOBEC-signature-DRM-filtered samples. NRTI and NNRTI DRMs were detected on almost equivalent levels in the non-filtered dataset, being present in 249/599 (41.6%) and 255/599 (42.6%) participants, respectively. Simultaneous presence of NRTI and NNRTI DRMs was detected in 183/599 (30.1%) participants before filtering out APOBEC-signature DRMs and in 17/599 (2.8%) after filtration. Major PI and InSTI DRMs were detected in 79/599 (13.2%) and 73/599 (12.2%) participants, respectively, considering the non-filtered dataset. The number of participants with DRMs in the

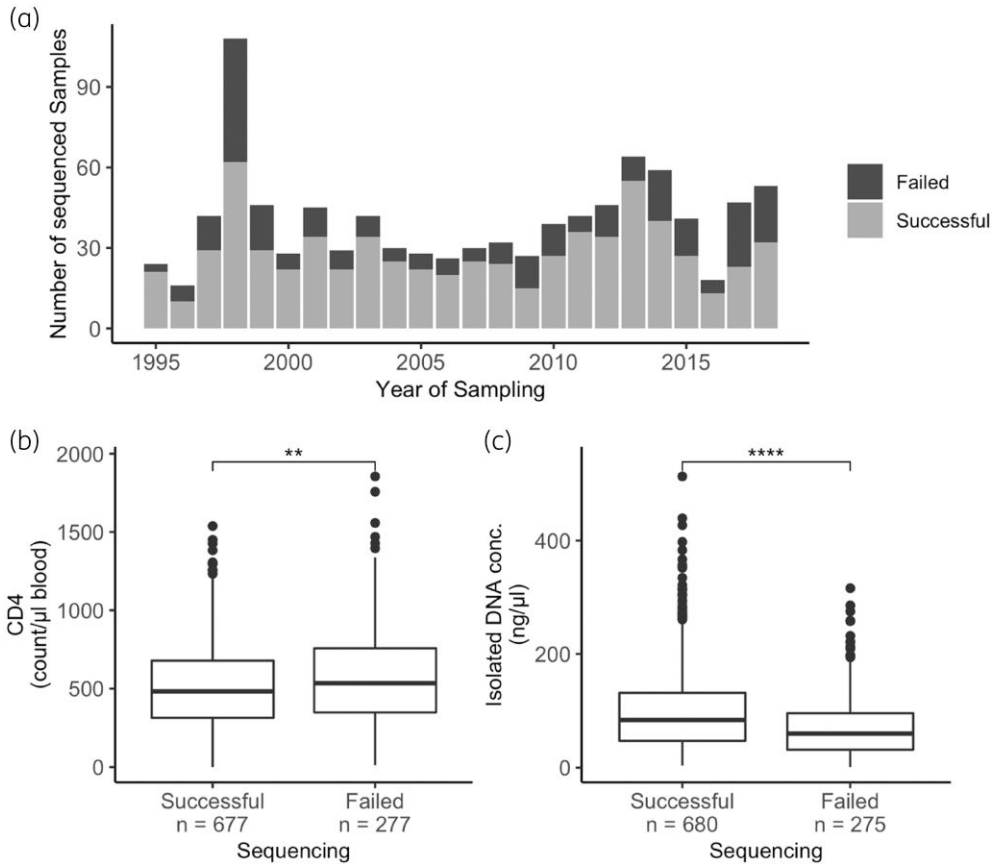

Figure 1. Characteristics of participants and blood specimens associated with successful or failed amplification and NGS of HIV-1 DNA. (a) Longitudinal distribution of the blood specimens stratified by outcome of amplification and NGS, showing resilience to temporal factors. (b) Significant difference in CD4+ T cell count, and (c) isolated DNA concentration from blood specimen. Mann–Whitney *U*: **, $P<0.01$; ****, $P<0.0001$.

latter drug classes dropped substantially after filtration of APOBEC-signature DRMs with 8/599 (1.3%) for PI DRMs and 3/599 (0.5%) for InSTI DRMs.

To determine differences between DRM carriers and non-carriers within the drug classes, we applied statistical tests on the aforementioned characteristics, with consistent stratification (Tables S3–S6). For NRTI DRMs, no differences were observed regarding age, sex, HIV-1 transmission group, year of sample collection (Figure 3a), therapy status, HIV-1 subtype, CD4+ T cell count and percentage of lymphocytes. After removing APOBEC-signature NRTI DRMs, the number of samples with detected NRTI DRMs dropped substantially from 249 to 99 (39.8%). Sex, HIV-1 transmission group and therapy status remained statistically insignificant. We observed a longer time on ART for NRTI DRM-carrying individuals in both analyses ($P<0.001$, $P<0.001$). Regarding ethnicities, both analyses revealed higher proportions of unknown ethnicity in NRTI DRM carriers while black ethnicity was found to be less prevalent in the APOBEC-filtered dataset for the NRTI DRM carriers ($P=0.003$, $P<0.001$; Table S3). Most NRTI DRM carriers, before and after filtering, started ART before 2000. Additionally, the participants with a detected NRTI DRM after applying the filtering algorithm showed more advanced age ($P=0.013$), earlier sample collection ($P<0.001$; Table S3 and Figure 3a), a significantly higher proportion of HIV-1 subtype B ($P=0.001$) and lower absolute and relative CD4+ T cell count ($P<0.001$; *t*-test, $P=0.001$).

Regarding NNRTIs, there was no difference between NNRTI DRM carriers and non-carriers in age, sex, ethnicities, HIV-1 transmission group, therapy status, HIV-1 subtype and the time of ART (Table S4). Nonetheless, the data revealed significant differences in the year of sample collection ($P=0.003$; Table S4, Figure 3a), which was later in NNRTI DRM carriers, and absolute ($P=0.014$) and relative CD4+ T cell count ($P=0.05$), which was higher in carriers. We again applied our filtering algorithm to investigate the impact of APOBEC3G/F-related G-to-A-mutations. Along with the number of samples with detected NNRTI DRMs decreasing by 76.5% (from 255 to 60), we observed fewer differences between the two groups. Additionally, for age, sex, ethnicities, HIV-1 transmission group, therapy status, HIV-1 subtype and the time of ART, as in the previous analysis, there was no difference between NNRTI carriers and non-carriers regarding absolute and relative CD4+ T cell count nor in viral load. Nevertheless, NNRTI carriers after filtering out APOBEC-signature DRMs showed a later year of sample collection (Table S4).

For PIs, we did not find significant differences between PI DRM carriers and non-carriers in age, sex, ethnicities, HIV-1 transmission group, year of sample collection (Figure 3a), therapy status, HIV-1 subtypes and absolute as well as relative CD4+ T cell count.

Table 2. List of all observed major DRMs in HIV-1 *pol* of the 331 participants with at least one detected DRM

Protease										
D30N	V32I	**M46I**	M46L	I54V	V82A	N88S	N88T	L90M		
53	2	55	2	2	3	1	2	4		
Reverse transcriptase										
M41L[a]	*E44A*	*E44D*	*A62V*	*K65R*	*D67E*[a]	***D67N***[a]	*D67T*[a]	*T69D*[b]	*T69I*[b]	*K70R*[a]	*L74V*
46	2	8	3	1	4	47	1	3	3	30	5
V75I	*V75M*	*F77L*	A98G	K101E	K103N	K103S	V106A	V108I	*F116Y*	E138A	**E138K**
8	1	1	5	2	13	1	1	7	1	22	85
E138Q	*Q151M*[b]	V179D	V179E	Y181C	***M184I***	*M184V*	G190A	**G190E**	**G190S**	*L210W*[a]	
1	1	8	3	1	181	19	1	19	2	17	
T215A[a]	*T215C*[a]	*T215D*[a]	*T215E*[a]	*T215F*[a]	*T215I*[a]	*T215L*[a]	*T215S*[a]	*T215Y*[a]	*K219E*[a]	*K219Q*[a]	
1	4	11	2	3	1	2	5	44	4	10	
K219R[a]		H221Y		**M230I**		Y318F				N348I	
1		3		210		1				3	
Integrase										
T66A	**G118R**	E138A	**E138K**	**G140S**	Q148H	N155T				
1	1	1	50	37	2	1				

Detected mutations encoded in the format: WT amino acid—position—mutated amino acid. The numbers below the mutations indicate the total number of detected mutations. In the RT region, mutations written in *italicized letters* represent NRTI , whereas mutations underlined represent NNRTI mutations in the RT gene. Highlighted mutations represent mutations caused by G-to-A changes in the nucleotide sequence, of which those in bold represent APOBEC-context mutations.

[a]TAMs, thymidine analogue mutations. Selected by ZDV and d4T; facilitate primer unblocking. Non-TAMs prevent NRTI incorporation.[24]

[b]Multi-drug resistance mutations. T69 insertions occur with TAMs. Q151M occurs with non-TAMs and the accessory mutations A62V, V75I, F77L and F116Y.[24]

Nonetheless, lower viral load ($P=0.01$) and longer time on ART ($P=0.005$, Table S5) was shown to have an impact on the prevalence of PI DRMs.

In Table S6 we reviewed the characteristics of InSTI DRM carriers. While statistical tests did not reveal significant difference in terms of age, ethnicity, HIV-1 transmission group, year of sample collection (Figure 3a), therapy status, HIV-1 subtype, CD4+ T cell count or duration of ART, InSTI DRM carriers were shown to differ from non-carriers regarding sex, with lower prevalence of women ($P=0.01$) and higher viral load ($P=0.012$). A high proportion of the InSTI DRM carriers were male (82.2%, 60/83).

Discussion

Our results demonstrate a method that successfully amplifies near full-length HIV-1 proviral DNA in a diverse population, encompassing a variety of HIV-1 subtypes. Furthermore, it allows analysis of DRMs affecting the most frequently used drugs in ART in a high proportion of the virologically suppressed and unsuppressed individuals using designated state-of-the-art bioinformatics tools.

Conventional GRT relies on circulating viral RNA and most often only targets specific genes in the HIV-1 *pol* using population-based Sanger sequencing.[11,36] Here we describe a method that uses the highly conserved long-terminal repeats (LTRs) as PCR primer binding sites, opening a gate to near full-length HIV-1 provirus sequencing information, particularly in a population of mostly ART-experienced and virologically suppressed individuals.[37] Applying this method, we successfully amplified and sequenced near full-length HIV-1 genomes in PBMCs from 681/962 (70.8%) HIV-1-infected individuals. We found several factors associated with successful sequencing of near full-length HIV-1 proviruses. One of them was the significantly lower CD4+ T cell count in individuals from whom amplification and sequencing of HIV-1 proviral DNA succeeded. A possible cause for failure of sequencing might therefore be the decay of the viral reservoir and the recovery of the CD4+ T cells to higher numbers over time under long-term successful ART.[38–40] Particularly in untreated patients, the viral genome is not only present in a state of integrated proviral DNA, but also to a substantial extent as non-integrated DNA, which can be amplified by this method, enhancing the yield of the PCR. Further, we observed a significantly lower concentration of isolated DNA from PBMC samples that failed sequence acquisition compared with those that were successful. Considering that our final protocol requires a fixed volume instead of a fixed amount of template DNA, this might have led to insufficient PCR input in a small number of cases. These points would be mandatory to consider for implementing our assay for diagnostic purposes. Another reason for unsuccessful amplification is the presence of deletions occurring naturally in proviral DNA that could lead to inability of primer binding.[30,31]

The commercialization of NGS platforms leads to rising availability with affordable application of NGS in daily clinical care and treatment.[36] This moves GRT of near full-length HIV-1 proviral DNA into focus within the field of drug resistance monitoring and it is necessary to understand the relevance of proviral DNA in drug resistance.[12–14] Coupled with MinVar, information about

Table 3. Characteristics of participants with and without detected DRMs in HIV-1 proviral DNA

		Not filtered			APOBEC-filtered			
Characteristic		DRM detected	No DRMs	*P* value	DRM detected	No DRMs	*P*	Test[a]
n		331	268		145	454		
Age (years), median (IQR)		41 (34–48)	40 (33–47)	0.264	42 [35, 48]	40 [33, 48]	0.201	mw
Sex, *n* (%)	Male	233 (70.4)	179 (66.8)	0.391	102 (70.3)	310 (68.3)	0.716	$\chi 2$
	Female	98 (29.6)	89 (33.2)		43 (29.7)	144 (31.7)		
Ethnicity, *n* (%)	White	208 (62.8)	182 (67.9)	0.019	86 (59.3)	304 (67)	<0.001	exact
	Black	79 (23.9)	56 (20.9)		29 (20)	106 (23.3)		
	Hispano-American	7 (2.1)	13 (4.9)		4 (2.8)	16 (3.5)		
	Asian	7 (2.1)	8 (3)		4 (2.8)	11 (2.4)		
	Unknown	29 (8.8)	9 (3.4)		22 (15.2)	16 (3.5)		
	Other	1 (0.3)	0 (0)		0 (0)	1 (0.2)		
HIV-1 transmission group, *n* (%)	MSM	137 (41.4)	106 (39.6)	0.663	58 (40)	185 (40.7)	0.598	exact
	HET	138 (41.7)	119 (44.4)		56 (38.6)	201 (44.3)		
	IDU	30 (9.1)	23 (8.6)		18 (12.4)	35 (7.7)		
	unclear (MSM, HET, IDU)	10 (3)	11 (4.1)		7 (4.8)	14 (3.1)		
	Blood products	0 (0)	1 (0.4)		0 (0)	1 (0.2)		
	Perinatal	5 (1.5)	2 (0.7)		2 (1.4)	5 (1.1)		
	Unknown	9 (2.7)	3 (1.1)		3 (2.1)	9 (2)		
	Other	2 (0.6)	3 (1.1)		1 (0.7)	4 (0.9)		
Year of sample collection, median (IQR)		2007 (2000–2013)	2007 (2000–2013)	0.651	2006 [1998, 2013]	2007 [2001, 2013]	0.14	mw
Therapy status, *n* (%)	On-ART	307 (94.5)	243 (92.0)	0.249	134 (93.7)	416 (93.3)	1	exact
	Pre-ART	18 (5.5)	21 (8.0)		9 (6.3)	30 (6.7)		
HIV-1 subtype, *n* (%)	B	224 (71.3)	174 (66.7)	0.239	103 (74.6)	295 (67.5)	0.138	exact
	Non-B	90 (28.7)	87 (33.3)		35 (25.4)	142 (32.5)		
CD4+ T cell count/mm^3 blood, median (IQR)		485.5 (313.2–676.5)	468 (313–675.5)	0.71	422.5 [252.8, 650]	495 [332, 681.5]	0.032	mw
CD4+ T cells (as % of lymphocytes), mean (SD)		25.8 (11.1)	25.4 (10.6)	0.681	23.9 (11.8)	26.2 (10.5)	0.026	*t*
Log_{10} of HIV-1 RNA copies/mL plasma, median (IQR)		0 (0–1.6)	0 (0–1.8)	0.582	0 [0, 2.3]	0 [0, 1.5]	0.02	mw
Time on ART (years), median (IQR)		2.9 (1.1–6.8)	1.9 (0.6–4.9)	0.001	3.6 [1.4, 8.1]	2.3 [0.7, 5.4]	<0.001	mw
PI DRM, *n* (%)	DRM detected	79 (23.9)	0 (0)		8 (5.5)	0 (0)		
	No DRM	252 (76.1)	268 (100)		137 (94.5)	454 (100)		
NRTI DRM, *n* (%)	DRM detected	249 (75.2)	0 (0)		99 (68.3)	0 (0)		
	No DRM	82 (24.8)	268 (100)		46 (31.7)	454 (100)		
NNRTI DRM, *n* (%)	DRM detected	255 (77)	0 (0)		60 (41.4)	0 (0)		
	No DRM	76 (23)	268 (100)		85 (58.6)	454 (100)		

Continued

Table 3. *Continued*

		Not filtered			APOBEC-filtered			
Characteristic		DRM detected	No DRMs	*P* value	DRM detected	No DRMs	*P*	Test[a]
InSTI DRM, *n* (%)	DRM detected	73 (22.1)	0 (0)		3 (2.1)	0 (0)		
	No DRM	258 (77.9)	268 (100)		142 (97.9)	454 (100)		
Year of ART initiation, *n* (%)	Before 2000	152 (45.9)	101 (37.7)	0.069	81 (55.9)	172 (37.9)	<0.001	exact
	2000–2006	72 (21.8)	82 (30.6)		21 (14.5)	133 (29.3)		
	After 2006	101 (30.5)	81 (30.2)		41 (28.3)	141 (31.1)		
	No ART	0 (0)	0 (0)		0 (0)	0 (0)		
	Unknown	6 (1.8)	4 (1.5)		2 (1.4)	8 (1.8)		
Year of infection (approx.), *n* (%)	1999–2005	56 (16.9)	69 (25.7)	0.05	13 (9)	112 (24.7)	<0.001	exact
	After 2005	88 (26.6)	65 (24.3)		33 (22.8)	120 (26.4)		
	Before 1999	168 (50.8)	116 (43.3)		87 (60)	197 (43.4)		
	Unknown	19 (5.7)	18 (6.7)		12 (8.3)	25 (5.5)		
Class resistance, *n* (%)	0	0 (0)	268 (100)		0 (0)	454 (100)		
	1	118 (35.6)			122 (84.1)			
	2	127 (38.4)			21 (14.5)			
	3	60 (18.1)			2 (1.4)			
	4	26 (7.9)			0 (0)			

MSM, men who have sex with men; HET, heterosexual; IDU, injection/intravenous drug use; SD, standard deviation; min, minimum; max, maximum; conc., concentration; mw, Mann–Whitney *U*/Wilcoxon rank-sum; χ^2, Pearson's chi-squared.

[a]Applied statistical test of independence between the stratified groups with *t* for two-sample *t*-test, mw for Mann–Whitney *U*/Wilcoxon rank-sum, and χ^2 for Pearson's chi-squared. Percentages refer to the respective category only.

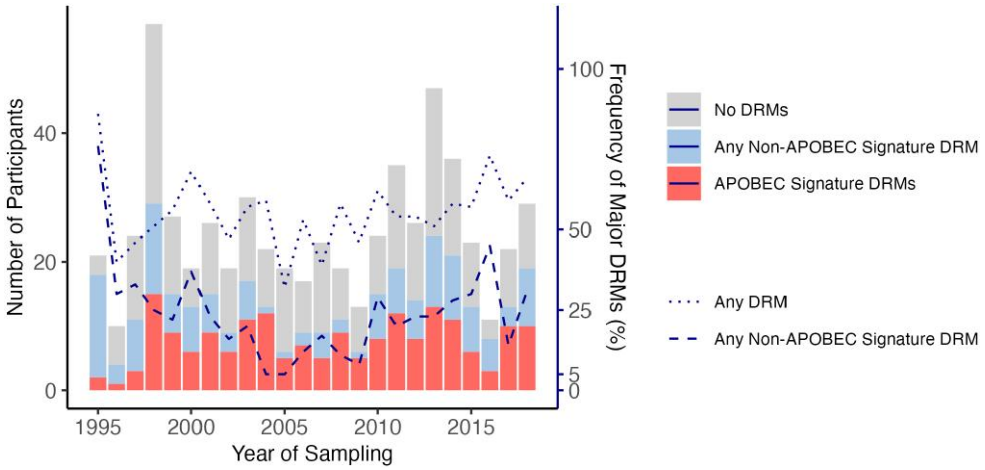

Figure 2. Longitudinal distribution of the 599 participant samples. Bar plot representation of the annual absolute number of participants in the observation period between 1995 and 2018 with any detected DRMs ($n = 331$, blue or red) and any DRMs that do not match the APOBEC-induced G-to-A signature ($n = 145$, blue). Patients without DRMs are shown in grey ($n = 268$). The frequencies of any DRMs (dark blue dotted line, right *y*-axis) as well as any non-APOBEC-signature DRMs (dashed line) are shown. This figure appears in colour in the online version of *JAC* and in black and white in the print version of *JAC*.

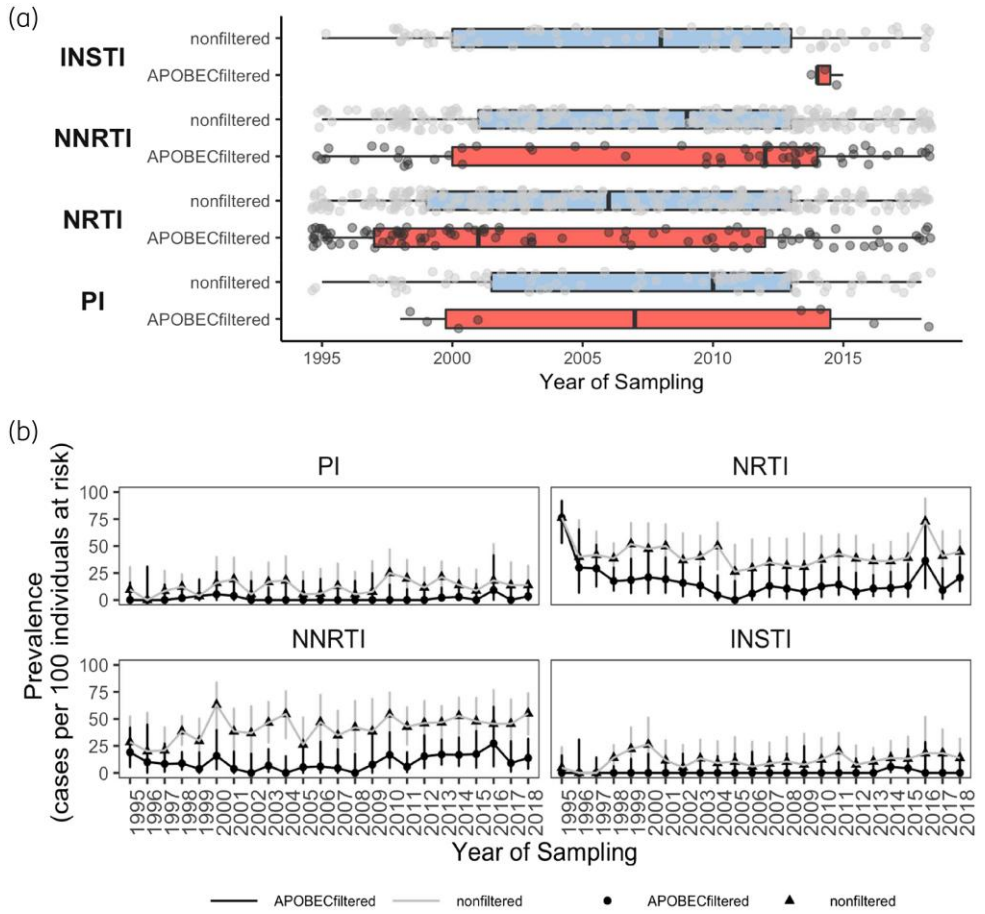

Figure 3. Prevalence of DRMs and differences between drug classes. Box plot representing the IQR and median of the samples with detected DRMs according to drug classes. The projected dots on each according box plot represent one sample in each category. Trends can be observed for the higher number of DRM-harbouring samples to PIs, NNRTIs and InSTIs in later years in comparison to DRMs to NRTIs (a). The prevalence for at least one DRM affecting the different drug classes has been estimated and visualized by dots (APOBEC-filtered) and triangles (non-filtered), which are flanked by error bars depicting the 95% CI, for each year for APOBEC-signature-DRM-filtered (solid line) and non-filtered samples (dotted line), respectively (b). This figure appears in colour in the online version of *JAC* and in black and white in the print version of *JAC*.

clinically relevant DRMs is presented in formatted reports that are easily interpreted by clinicians for routine use.[22] In addition, generated NGS information could not only be used for treatment adjustment in individuals but also for studies focusing on the understanding of transmission patterns and networks, intra-host evolution and viral reservoirs—the main hurdles to HIV cure.[16]

Mutations in the HIV-1 genome can usually be traced back to its own error-prone viral RT.[41] However, there are known cellular factors contributing to the induction of mutations. APOBEC3G/F-mediated G-to-A mutation is a known cellular host defence mechanism in retroviral infections, also known under the term of hypermutation.[42] The most abundantly found DRMs in our cohort are all listed in the APOBEC-signature mutations list of the Stanford HIV-1 database.[24] Since the mutagenic step in this mechanism is restricted to the cDNA, proviral DNA is more prone to harbouring these mutations compared with plasma RNA.[43] High rates of hypermutations have been shown to cause restriction of HIV-1 replication by introduction of stop codons or frameshift mutations, thus leading to defective proviruses and consequently neglectable for analysis of drug resistance.[17,42] In single-cell based assays, GRT is possible due to discrimination of replication-competent viruses from potentially hypermutated and defective viruses; however, in an NGS-based bulk PCR approach, such as presented in this study, this discrimination would have to be performed on the read level, so as to avoid overcorrection by exclusion on a per-mutation basis.

The results of our study showed significant variations in the prevalence of DRMs among the different commonly used drug classes in ART. The accumulation of DRMs through longer drug exposure seems intuitive, especially when patients were exposed to historic ART regimens with low potency and/or low genetic barrier to resistance or high toxicity, which can result in adherence problems and insufficient viral suppression.[44] Markedly, NRTI and NNRTI DRMs, the two most commonly used drugs in first-line ART throughout the past two decades, were detected in high abundance, with a large proportion of DRMs matching the APOBEC signature. Alongside the widespread use in clinical care, the preferential target of the RT by APOBEC might also contribute to this finding.[45] Filtering out the APOBEC-signature DRM provided a more intuitive interpretation of the results regarding NRTIs and NNRTIs. In particular, the higher CD4+ T cell count in NNRTI DRM carriers was difficult to comprehend. For NRTI DRM carriers, the opposite effect was observed, indicating potential immune suppression through viral replication.

The retrospective nature of this study delivers a range of different variables to analyse. However, the conjointly harboured heterogeneity has limitations in the interpretation of the data. The relatively small sample size when looking at specific aspects, e.g. ART status and viral load, limits the statistical significance of tests and models. Further, having access to previous GRT data of the participants would have been of interest to evaluate differences between HIV-1 RNA- and DNA-oriented approaches in the same individuals. The interpretation of APOBEC-signature DRM in the context of proviral DNA poses a major challenge in studies using HIV-1 DNA as a target. This challenge applies also to our study. Our method to filter out APOBEC-signature DRM addresses the need for elaborated bioinformatic tools to analyse proviral HIV-1 sequences.

Conclusions

To conclude, our results represent a reliable GRT method that successfully amplifies, sequences and analyses near full-length HIV-1 genomes, using HIV-1 proviral DNA from PBMCs as a template in the majority of a demographically and clinically diverse population of ART-naive, ART-experienced and virologically suppressed HIV-1-infected individuals. Although uncertainties about the interpretation of NGS data based on HIV-1 proviral DNA exist, emerging data from sequences like ours offer potential insights for a more comprehensive analysis, best combined with designated bioinformatics tools, to elicit clinical significance of mutational changes in HIV-1 proviral DNA. Particularly in long-term suppressed ART patients, GRT in proviral DNA might offer guidance to clinicians considering treatment simplification in the future.[46] Future large-scale studies, most optimally with prospective or longitudinal character, are needed, firstly to determine the differences between the predictive value of HIV-1 RNA and DNA sequences, secondly to understand the importance of possible resulting differences for viral behaviour, and thirdly to investigate the clinical implications and applications of proviral DNA GRT.

Acknowledgements

We thank the participants of the SHCS, the study nurses, physicians, data managers and the administrative assistants. We thank Gabriela Ziltener for technical assistance.

Members of the SHCS

A. Anagnostopoulos, M. Battegay, E. Bernasconi, J. Böni, D. L. Braun, H. C. Bucher, A. Calmy, M. Cavassini, A. Ciuffi, G. Dollenmaier, M. Egger, L. Elzi, J. Fehr, J. Fellay, H. Furrer, C. A. Fux, H. Günthard (President of the SHCS), D. Haerry (deputy of 'Positive Council'), B. Hasse, H. H. Hirsch, M. Hoffmann, I. Hösli, M. Huber, C. R. Kahlert (Chairman of the Mother and Child Substudy), L. Kaiser, O. Keiser, T. Klimkait, R. D. Kouyos, H. Kovari, K. Kusejko (Head of Data Centre), B. Ledergerber, G. Martinetti, B. Martinez de Tejada, C. Marzolini, K. J. Metzner, N. Müller, D. Nicca, P. Paioni, G. Pantaleo, M. Perreau, A. Rauch (Chairman of the Scientific Board), C. Rudin, P. Schmid, R. Speck, M. Stöckle (Chairman of the Clinical and Laboratory Committee), P. Tarr, A. Trkola, P. Vernazza, G. Wandeler, R. Weber and S. Yerly.

Funding

This study has been financed within the framework of the Swiss HIV Cohort Study, supported by the Swiss National Science Foundation (grant #177499, #201369 and grant #179571 to H.F.G.) and by the Yvonne Jacob Foundation (to H.F.G.).

Transparency declarations

H.F.G. has received unrestricted research grants from Gilead Sciences and fees for data and safety monitoring board membership, for advisory board and consulting activities from Gilead Sciences, Janssen, Johnson and Johnson, ViiV, Novartis and Merck and a travel grant from Gilead. K.J.M. has received travel grants and honoraria from Gilead Sciences, Roche Diagnostics, GlaxoSmithKline, Merck Sharp & Dohme, Bristol-Myers Squibb, ViiV and Abbott; and the University of Zurich

received research grants from Gilead Science, Novartis, Roche, and Merck Sharp & Dohme for studies that K.J.M. serves as principal investigator for, and advisory board honoraria from Gilead Sciences and ViiV. All other authors declare no conflict of interest.

Author contributions

B.J., C.D.S., H.F.G. and K.J.M. developed the study protocol. B.J., C.D.S., L.J., K.N. and C.L. performed the experiments. B.J. and C.D.S. analysed the data. M.H., M.Z., R.D.K. and K.J.M. supported the data analyses. A.C. and M.C. contributed cell samples and clinical data. B.J. drafted the manuscript, supported by K.J.M. All authors approved the final version of the manuscript.

Supplementary data

Figure S1 and Tables S1 to S6 are available as Supplementary data at *JAC* Online.

References

1 The Antiretroviral Therapy Cohort Collaboration. Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet* 2008; **372**: 293–9. https://doi.org/10.1016/S0140-6736(08)61113-7

2 Gueler A, Moser A, Calmy A *et al*. Life expectancy in HIV-positive persons in Switzerland: matched comparison with general population. *AIDS* 2017; **31**: 427–36. https://doi.org/10.1097/QAD.0000000000001335

3 Vercauteren J, Wensing AMJ, Van De Vijver DAMC *et al*. Transmission of drug-resistant HIV-1 is stabilizing in Europe. *J Infect Dis* 2009; **200**: 1503–8. https://doi.org/10.1086/644505

4 Scherrer AU, von Wyl V, Yang W-L *et al*. Emergence of acquired HIV-1 drug resistance almost stopped in Switzerland: a 15-year prospective cohort analysis. *Clin Infect Dis* 2016; **62**: 1310–7. https://doi.org/10.1093/cid/ciw128

5 von Wyl V, Yerly S, Böni J, *et al*. Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. *Arch Intern Med* 2007; **167**: 1782–90. https://doi.org/10.1001/archinte.167.16.1782

6 Davidson I, Beardsell H, Smith B *et al*. The frequency and reasons for antiretroviral switching with specific antiretroviral associations: the SWITCH study. *Antiviral Res* 2010; **86**: 227–9. https://doi.org/10.1016/j.antiviral.2010.03.001

7 Günthard HF, Calvez V, Paredes R *et al*. Human immunodeficiency virus drug resistance: 2018 recommendations of the International Antiviral Society–USA panel. *Clin Infect Dis* 2019; **68**: 177–87. https://doi.org/10.1093/cid/ciy463

8 WHO. Global Action Plan on HIV Drug Resistance 2017-2021: 2017 Progress Report - HIV Drug Resistance. 2017. https://apps.who.int/iris/bitstream/handle/10665/255883/9789241512848-eng.pdf?sequence=1&isAllowed=y.

9 Vandamme A-M, Camacho RJ, Ceccherini-Silberstein F *et al*. European recommendations for the clinical use of HIV drug resistance testing: 2011 update. *AIDS Rev* 2011; **13**: 77–108.

10 Santoro MM, Fabeni L, Armenia D *et al*. Reliability and clinical relevance of the HIV-1 drug resistance test in patients with low viremia levels. *Clin Infect Dis* 2014; **58**: 1156–64. https://doi.org/10.1093/cid/ciu020

11 Clutter DS, Jordan MR, Bertagnolio S *et al*. HIV-1 drug resistance and resistance testing. *Infect Genet Evol* 2016; **46**: 292–307. https://doi.org/10.1016/j.meegid.2016.08.031

12 Novitsky V, Zahralban-Steele M, McLane MF *et al*. Long-range HIV genotyping using viral RNA and proviral DNA for analysis of HIV drug resistance and HIV clustering. *J Clin Microbiol* 2015; **53**: 2581–92. https://doi.org/10.1128/JCM.00756-15

13 Meybeck A, Alidjinou EK, Huleux T *et al*. Virological outcome after choice of antiretroviral regimen guided by proviral HIV-1 DNA genotyping in a real-life cohort of HIV-infected patients. *AIDS Patient Care STDS* 2020; **34**: 51–8. https://doi.org/10.1089/apc.2019.0198

14 Zaccarelli M, Santoro MM, Armenia D *et al*. Genotypic resistance test in proviral DNA can identify resistance mutations never detected in historical genotypic test in patients with low level or undetectable HIV-RNA. *J Clin Virol* 2016; **82**: 94–100. https://doi.org/10.1016/j.jcv.2016.07.007

15 Allavena C, Rodallec A, Leplat A *et al*. Interest of proviral HIV-1 DNA genotypic resistance testing in virologically suppressed patients candidate for maintenance therapy. *J Virol Methods* 2018; **251**: 106–10. https://doi.org/10.1016/j.jviromet.2017.10.016

16 Metzner KJ. Technologies for HIV-1 drug resistance testing: inventory and needs. *Curr Opin HIV AIDS* 2022; **17**: 222–8. https://doi.org/10.1097/COH.0000000000000737

17 Carolyn C, Daniele A, Charles W *et al*. Genotypic resistance testing of HIV-1 DNA in peripheral blood mononuclear cells. *Clin Microbiol Rev* 2022; **35**: e00052-22. https://doi.org/10.1128/cmr.00052-22

18 Scherrer AU, Traytel A, Braun DL *et al*. Cohort profile update: the Swiss HIV cohort study (SHCS). *Int J Epidemiol* 2022; **51**: 33–4. https://doi.org/10.1093/ije/dyab141

19 Gall A, Ferns B, Morris C *et al*. Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *J Clin Microbiol* 2012; **50**: 3838–44. https://doi.org/10.1128/JCM.01516-12

20 Banin AN, Tuen M, Bimela JS *et al*. Development of a versatile, near full genome amplification and sequencing approach for a broad variety of HIV-1 group M variants. *Viruses* 2019; **11**: 317. https://doi.org/10.3390/v11040317

21 Di GF, Töpfer A, Rey M *et al*. Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. *Nucleic Acids Res* 2014; **42**: e115. https://doi.org/10.1093/nar/gku537

22 Huber M, Metzner KJ, Geissberger FD *et al*. MinVar: a rapid and versatile tool for HIV-1 drug resistance genotyping by deep sequencing. *J Virol Methods* 2017; **240**: 7–13. https://doi.org/10.1016/j.jviromet.2016.11.008

23 Foley BT, Korber BTM, Leitner TK *et al*. HIV Sequence Compendium 2018. 2018. https://www.osti.gov/biblio/1458915.

24 Rhee S-Y, Gonzales MJ, Kantor R *et al*. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 2003; **31**: 298–303. https://doi.org/10.1093/nar/gkg100

25 R Core Team. R: A Language and Environment for Statistical Computing. 2020.

26 Wickham H, Averick M, Bryan J *et al*. Welcome to the Tidyverse. *J Open Source Softw* 2019; **4**: 1686. https://doi.org/10.21105/joss.01686

27 Yoshida K, Bartel A, Chipman JJ *et al*. tableone: Create 'Table 1' to Describe Baseline Characteristics with or without Propensity Score Weights. 2020. https://rdrr.io/cran/tableone/.

28 Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, 2016.

29 Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. 2020. https://rpkgs.datanovia.com/ggpubr/.

30 Bruner KM, Murray AJ, Pollack RA *et al*. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* 2016; **22**: 1043–9. https://doi.org/10.1038/nm.4156

31 Ho Y-C, Shan L, Hosmane NN *et al*. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* 2013; **155**: 540–51. https://doi.org/10.1016/j.cell.2013.09.020

32 Wensing AM, Calvez V, Ceccherini-Silberstein F *et al*. 2019 update of the drug resistance mutations in HIV-1. *Top Antivir Med* 2019; **27**: 111–21.

33 Shafer RW. Rationale and uses of a public HIV drug-resistance database. *J Infect Dis* 2006; **194**: S51–8. https://doi.org/10.1086/505356

34 Rose PP, Korber BT. Detecting hypermutations in viral sequences with an emphasis on G → A hypermutation. *Bioinformatics* 2000; **16**: 400–1. https://doi.org/10.1093/bioinformatics/16.4.400

35 Stanford University HIV Drug Resistance Database. HIV-1 Signature APOBEC Mutations. 2021: 222–8. https://hivdb.stanford.edu/page/apobecs/.

36 Van Laethem K, Theys K, Vandamme A-M. HIV-1 genotypic drug resistance testing: digging deep, reaching wide? *Curr Opin Virol* 2015; **14**: 16–23. https://doi.org/10.1016/j.coviro.2015.06.001

37 UNAIDS. UNAIDS Global Factsheet 2020. 2020. https://aidsinfo.unaids.org.

38 Wan C, Bachmann N, Mitov V *et al*. Heritability of the HIV-1 reservoir size and decay under long-term suppressive ART. *Nat Commun* 2020; **11**: 5542. https://doi.org/10.1038/s41467-020-19198-7

39 Bachmann N, von Siebenthal C, Vongrad V *et al*. Determinants of HIV-1 reservoir size and long-term dynamics during suppressive ART. *Nat Commun* 2019; **10**: 3193. https://doi.org/10.1038/s41467-019-10884-9

40 Pinzone MR, VanBelzen DJ, Weissman S *et al*. Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat Commun* 2019; **10**: 728. https://doi.org/10.1038/s41467-019-08431-7

41 Yeo JY, Goh G-R, Su CT-T *et al*. The determination of HIV-1 RT mutation rate, its possible allosteric effects, and its implications on drug resistance. *Viruses* 2020; **12**: 297. https://doi.org/10.3390/v12030297

42 Harris RS, Liddament MT. Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* 2004; **4**: 868–77. https://doi.org/10.1038/nri1489

43 Noguera-Julian M, Cozzi-Lepri A, Di Giallonardo F *et al*. Contribution of APOBEC3G/F activity to the development of low-abundance drug-resistant human immunodeficiency virus type 1 variants. *Clin Microbiol Infect* 2016; **22**: 191–200. https://doi.org/10.1016/j.cmi.2015.10.004

44 Tang MW, Shafer RW. HIV-1 antiretroviral resistance. *Drugs* 2012; **72**: e1–25. https://doi.org/10.2165/11633630-000000000-00000

45 Suspène R, Rusniok C, Vartanian J-P *et al*. Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. *Nucleic Acids Res* 2006; **34**: 4677–84. https://doi.org/10.1093/nar/gkl555

46 Gandhi RT, Bedimo R, Hoy JF *et al*. Antiretroviral drugs for treatment and prevention of HIV infection in adults: 2022 recommendations of the International Antiviral Society–USA panel. *JAMA* 2023; **329**: 63–84. https://doi.org/10.1001/jama.2022.22246